

## **Supplementary information**

### **Sequence diversity of the Rh blood group system in Basques**

André Flores-Bello<sup>1</sup>, David Mas-Ponte<sup>2</sup>, Miruna E. Rosu<sup>3</sup>, Elena Bosch<sup>1</sup>, Francesc Calafell<sup>1</sup>, David Comas<sup>1</sup>

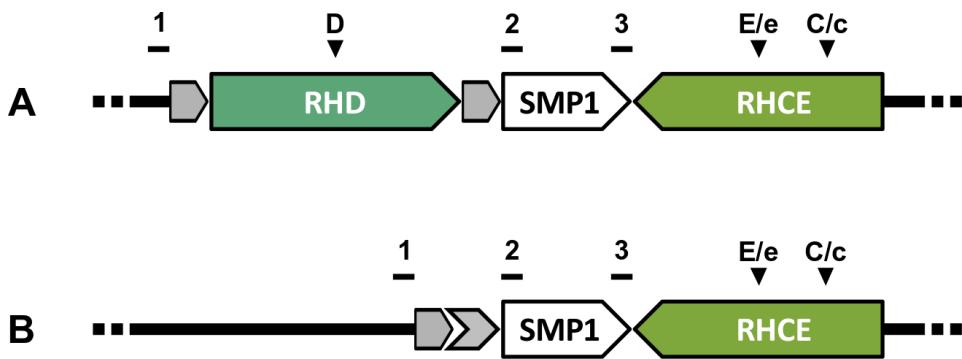
<sup>1</sup>Institute of Evolutionary Biology (CSIC – UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain.

<sup>2</sup>Institute for Research in Biomedicine Barcelona, Barcelona, Spain.

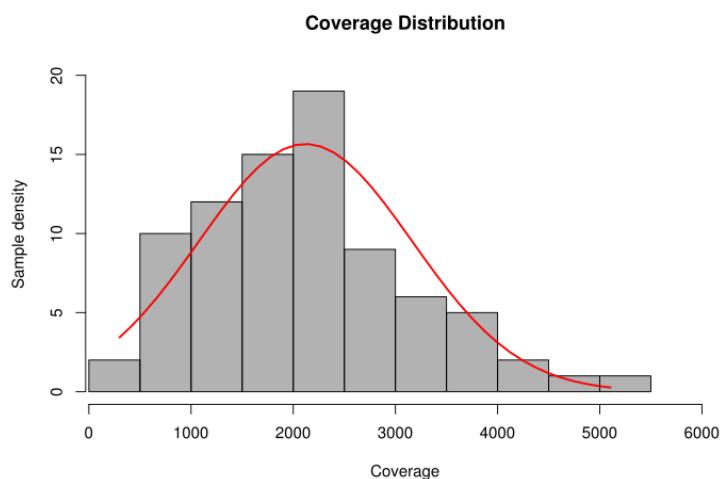
<sup>3</sup>Department of Viroscience, Postgraduate School Molecular Medicine, Erasmus MC, Rotterdam, The Netherlands.

Corresponding author: Dr. David Comas Martínez, Institute of Evolutionary Biology (CSIC – UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08003, Barcelona, Catalonia, Spain. Tel.: +933160841; E-mail address: david.comas@upf.edu

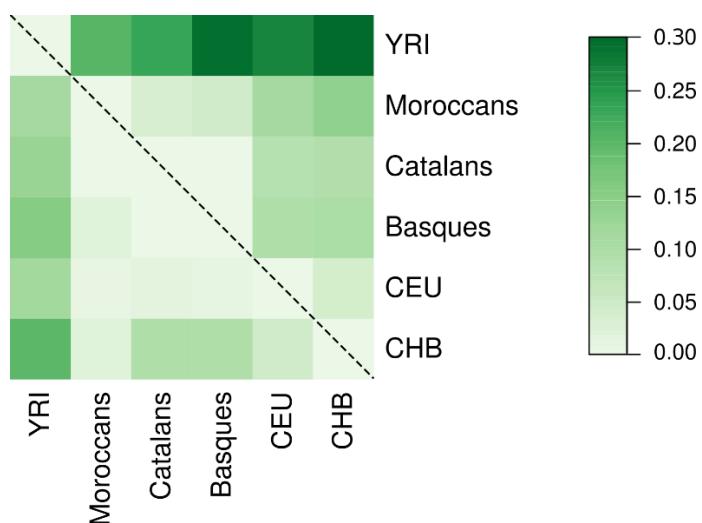
**Figure S1. Schematic genomic landscape of the Rh system.** Genes and *Rhesus boxes*, segmental duplications with 99% sequence identity flanking the *RHD* gene that contain the RHD deletion breakpoint, are represented by arrows defining their position and orientation<sup>1</sup>. D, E/e and C/c antigens are represented by a triangle above the *RHD* and *RHCE* genes, respectively. 1, 2 and 3 are the studied flanking regions: *RHD* upstream, *RHD* downstream and *RHCE* downstream, respectively. A: genomic landscape in an RhD positive chromosome; B: genomic landscape in an RhD negative chromosome given by the deletion of the *RHD* gene.



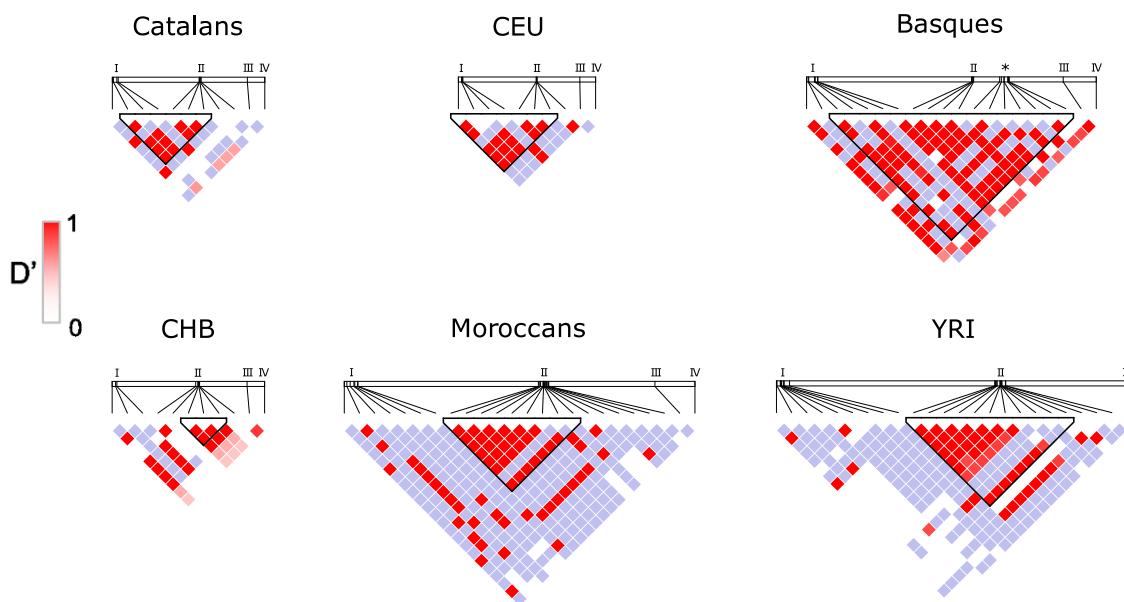
**Figure S2. Sequencing coverage histogram.** Sequencing coverage mean was calculated for each sample with the Genome Analysis ToolKit software<sup>2</sup>.



**Figure S3. Heatmap from pairwise  $F_{ST}$  comparisons between populations.** RHD upstream flanking region is represented above the diagonal while the RHD downstream is shown below the diagonal.

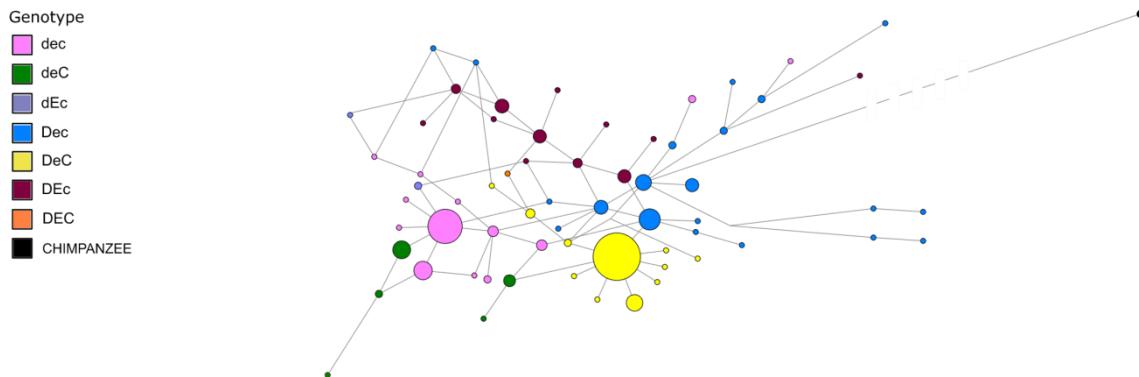
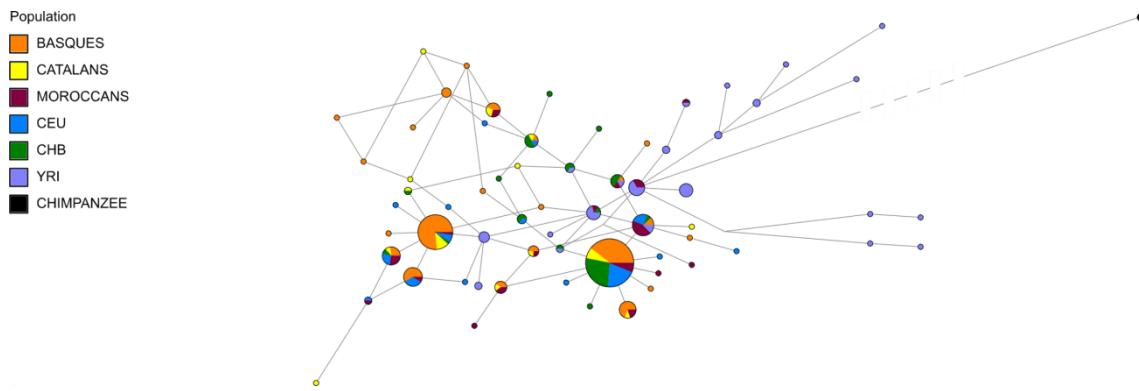


**Figure S4. Linkage disequilibrium graphics for the three regions computed together in each population.** Performed by Haploview v4.2<sup>3</sup>. Blocks were estimated by the “Confidence Intervals” algorithm defined by Gabriel et al. (2002)<sup>4</sup>. Hardy-Weinberg p-value and minimum Minor Allele Frequency cut-off were set at 0.0001 and 0.05 respectively. The bar above each graphic represents the studied regions from chromosome 1, and the lines are those SNPs that passed the cut-off. A different symbol has been used to represent every region. **I:** RHD upstream; **II:** RHD downstream; **III:** Ee allele; **IV:** Cc allele; **\***: RHCE downstream, whose SNPs only passed the cut-off for Basque population. No Basque specific block was detected in the analysis. Instead, all the populations show a LD block in the RHD downstream region. In Europeans, this block extends up to the RHD upstream region.

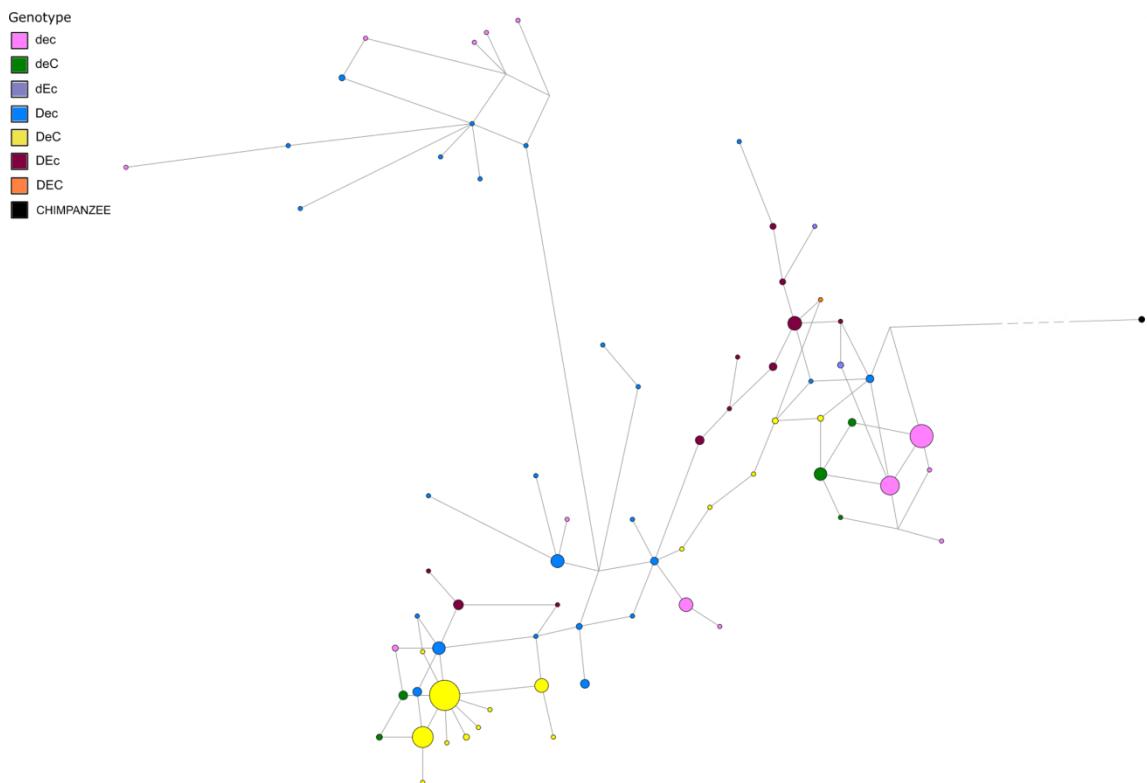
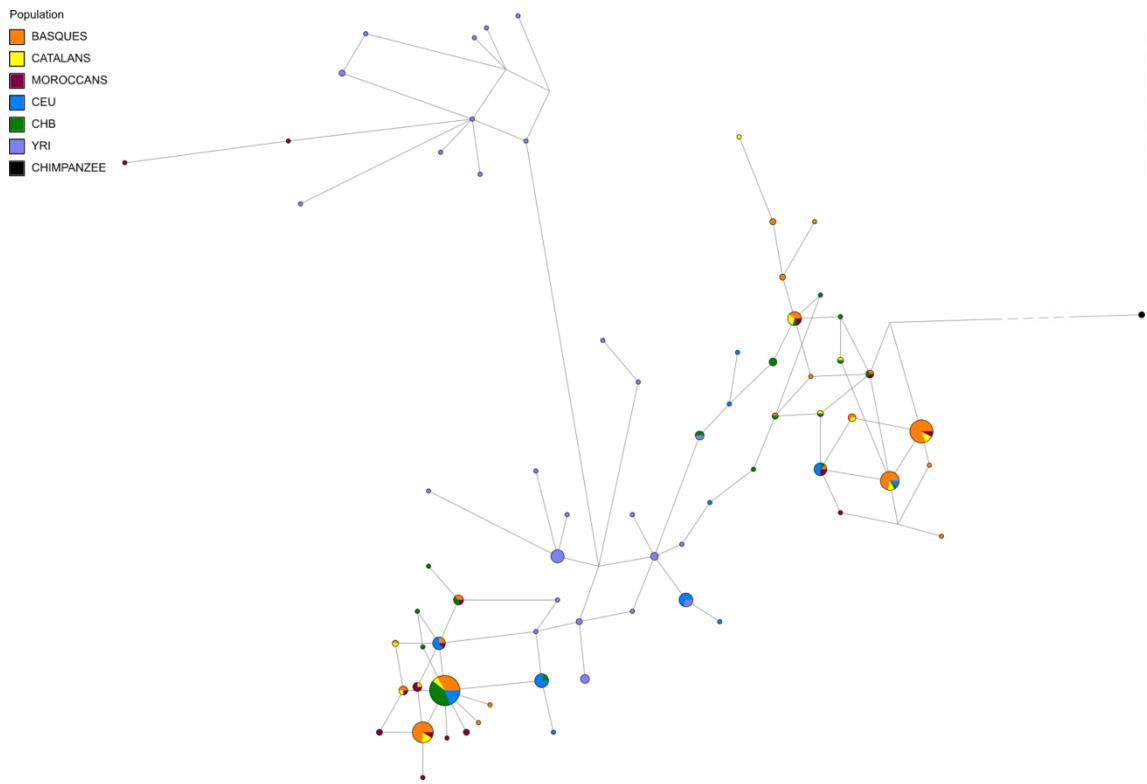


**Figure S5. Network analysis of haplotypes from both RHD upstream (A) and RHD downstream (B) flanking region separately. Coloured by population and the three RHD/RHCE polymorphisms haplotypes.**

**(A) RHD upstream flanking region.**



**(B) RHD downstream flanking region.**



**Table S1. Features of primers used in the amplification of the RHD/RHCE flanking regions.**

Region	Primer	Sequence (5' to 3')	Tm (°C)
RHD upstream	Forward	AAATAGAAGGTGCATTGAGCAAA	60.1
RHD upstream	Reverse	TTAGCTTACTGGATGACCACCAT	59.9
RHD downstream	Forward	AGACGCCTTGTGTTGGTG	59.3
RHD downstream	Reverse	GAACTGCAGCATTCAAGAAGAAT	59.9
RHCE downstream	Forward	TCTTCCCTCTTGTGCCTACG	60.4
RHCE downstream	Reverse	TCTTGTCTACATCAGCAAAGCA	58.7

**Table S2A. D/d, E/e and C/c genotypes for our Basque, Catalan and Moroccan samples and those CEU, CHB and YRI included from Perry et al. (2012)<sup>5</sup>.**

Population	Sample Name	RHD copies	C/c genotype	E/e genotype
Basque Country	B4	1	cc	Ee
Basque Country	B5	1	Cc	ee
Basque Country	B6	2	Cc	Ee
Basque Country	B7	1	Cc	ee
Basque Country	B10	1	Cc	ee
Basque Country	B12	1	cc	ee
Basque Country	B14	0	cc	ee
Basque Country	B17	0	cc	ee
Basque Country	B18	0	cc	ee
Basque Country	B19	0	cc	ee
Basque Country	B20	1	Cc	ee
Basque Country	B22	2	CC	ee
Basque Country	B23	1	Cc	ee
Basque Country	B25	1	Cc	ee
Basque Country	B26	1	Cc	ee
Basque Country	B27	1	CC	ee
Basque Country	B28	1	Cc	ee
Basque Country	B29	1	Cc	ee
Basque Country	B30	1	Cc	ee
Basque Country	B31	2	CC	ee
Basque Country	B32	2	cc	Ee
Basque Country	B41	1	cc	Ee
Basque Country	B43	2	CC	ee
Basque Country	B44	1	Cc	ee
Basque Country	B62	0	cc	ee
Basque Country	B64	1	Cc	ee
Basque Country	B65	0	cc	ee
Basque Country	B66	0	cc	ee
Basque Country	B67	1	Cc	ee
Basque Country	B69	0	cc	ee
Basque Country	B70	2	CC	ee
Basque Country	B73	2	CC	ee
Basque Country	B74	0	cc	ee
Basque Country	B76	0	cc	ee
Basque Country	B78	1	Cc	ee
Basque Country	B79	2	Cc	Ee

Basque Country	B80	0	cc	ee
Basque Country	B81	0	cc	ee
Basque Country	B83	2	Cc	Ee
Basque Country	B85	2	cc	Ee
Basque Country	B90	2	CC	ee
Basque Country	B92	2	CC	ee
Basque Country	B94	1	cc	Ee
Basque Country	B95	1	Cc	Ee
Basque Country	B96	2	CC	ee
Basque Country	B99	2	CC	ee
Basque Country	B100	2	CC	ee
Basque Country	B103	0	cc	ee
Basque Country	B104	1	cc	Ee
Basque Country	B105	1	cc	Ee
Basque Country	B106	1	CC	ee
Basque Country	B107	2	CC	ee
Basque Country	B109	1	CC	ee
Catalonia	C10	2	Cc	Ee
Catalonia	C16	1	cc	Ee
Catalonia	C18	1	cc	Ee
Catalonia	C27	1	cc	Ee
Catalonia	C39	0	cc	ee
Catalonia	C48	2	CC	ee
Catalonia	C51	0	Cc	ee
Catalonia	C56	2	Cc	ee
Catalonia	C57	2	CC	ee
Catalonia	C59	1	CC	ee
Catalonia	C66	1	Cc	ee
Catalonia	C82	1	cc	Ee
Morocco	M11	2	Cc	ee
Morocco	M12	2	cc	ee
Morocco	M14	2	cc	ee
Morocco	M15	2	Cc	ee
Morocco	M16	2	Cc	ee
Morocco	M17	1	CC	ee
Morocco	M18	1	Cc	ee
Morocco	M19	1	Cc	ee
Morocco	M20	2	CC	ee
Morocco	M25	0	Cc	ee
Morocco	M32	1	CC	ee
Morocco	M33	1	Cc	ee
Morocco	M40	1	Cc	ee
Morocco	M44	1	Cc	Ee
Morocco	R5	2	Cc	ee
Morocco	R47	1	cc	Ee
Morocco	R50	1	cc	Ee
CEU	CU1	1	Cc	ee
CEU	CU2	2	CC	ee
CEU	CU3	1	Cc	ee
CEU	CU4	1	Cc	ee
CEU	CU5	2	CC	ee

CEU	CU6	2	CC	ee
CEU	CU7	2	CC	ee
CEU	CU8	1	Cc	ee
CEU	CU9	1	Cc	ee
CEU	CU10	1	cc	ee
CEU	CU11	1	Cc	ee
CEU	CU12	1	Cc	ee
CEU	CU13	1	Cc	ee
CEU	CU14	1	Cc	ee
CEU	CU15	1	cc	Ee
CEU	CU16	0	cc	ee
CEU	CU17	1	Cc	ee
CEU	CU18	2	Cc	Ee
CEU	CU19	2	CC	ee
CEU	CU20	2	CC	ee
CHB	CH1	1	CC	ee
CHB	CH2	2	CC	ee
CHB	CH3	2	CC	ee
CHB	CH4	1	cc	Ee
CHB	CH5	2	CC	ee
CHB	CH6	2	CC	Ee
CHB	CH7	2	Cc	Ee
CHB	CH8	2	Cc	Ee
CHB	CH9	2	Cc	Ee
CHB	CH10	2	Cc	Ee
CHB	CH11	2	Cc	Ee
CHB	CH12	2	CC	ee
CHB	CH13	2	Cc	Ee
CHB	CH14	2	CC	ee
CHB	CH15	2	CC	ee
CHB	CH16	2	Cc	Ee
CHB	CH17	2	CC	ee
CHB	CH18	2	Cc	Ee
CHB	CH19	1	cc	ee
CHB	CH20	2	cc	EE
CHB	CH21	2	CC	ee
YRI	YR1	2	cc	Ee
YRI	YR2	2	cc	Ee
YRI	YR3	2	cc	ee
YRI	YR4	1	cc	ee
YRI	YR5	2	cc	ee
YRI	YR6	2	cc	ee
YRI	YR7	1	cc	ee
YRI	YR8	2	cc	ee
YRI	YR9	1	cc	ee
YRI	YR10	2	cc	ee
YRI	YR11	2	cc	ee
YRI	YR12	2	cc	ee
YRI	YR13	2	cc	ee
YRI	YR14	2	cc	ee
YRI	YR15	1	cc	ee

YRI	YR16	2	cc	ee
YRI	YR17	1	cc	ee
YRI	YR18	1	cc	ee
YRI	YR19	2	cc	ee
YRI	YR20	2	cc	ee
YRI	YR21	1	cc	Ee
YRI	YR22	1	cc	ee

**Table S2B. D/d, E/e and C/c genotypes for our Basque, Catalan and Moroccan samples and those CEU, CHB and YRI included from Perry et al. (2012)<sup>5</sup>.**

	DD	Dd	dd	EE	Ee	ee	CC	Cc	cc
Basques	30.2	45.3	24.5	0.0	20.8	79.2	26.4	34.0	39.6
Catalans	33.3	50.0	16.7	0.0	41.7	58.3	25.0	33.3	41.7
Moroccans	41.2	52.9	5.9	0.0	17.6	82.4	17.6	58.8	23.5
CEU	35.0	60.0	5.0	0.0	10.0	90.0	30.0	55.0	15.0
CHB	85.7	14.3	0.0	4.8	47.6	47.6	47.6	38.1	14.3
YRI	63.6	36.4	0.0	0.0	13.6	86.4	0.0	0.0	100.0

**Table S3. Frequencies of the RHD/RHCE variants and their haplotypes.** N represents the number of available haplotypes in each population.

N	Dd/Ee/Cc Haplotypes							Dd	Ee	Cc			
	dec	deC	dEc	Dec	DeC	DEc	DEC						
Basques	106.00	0.42	0.05	0.01	0.05	0.39	0.09	0.00	0.53	0.47	0.90	0.10	0.57 0.43
Catalans	24.00	0.29	0.08	0.04	0.08	0.33	0.17	0.00	0.58	0.42	0.79	0.21	0.58 0.42
Moroccans	34.00	0.12	0.21	0.00	0.32	0.26	0.09	0.00	0.68	0.32	0.91	0.09	0.53 0.47
CEU	40.00	0.25	0.10	0.00	0.13	0.48	0.05	0.00	0.65	0.35	0.95	0.05	0.43 0.58
CHB	42.00	0.02	0.02	0.02	0.05	0.62	0.24	0.02	0.93	0.07	0.71	0.29	0.33 0.67
YRI	44.00	0.18	0.00	0.00	0.73	0.02	0.07	0.00	0.82	0.18	0.93	0.07	0.98 0.02

**Table S4. RHD/RHCE flanking regions haplotypes table by individual.** All haplotypes found by individual sample after phasing the data are represented in this table. SNP coordinates are based on the GRCh37/hg19 assembly. Moreover, if available, SNP IDs are showed in the second row.

**Table S5A. Summary table of haplotypes from RHDup region represented by population.**

**Table S5B. Summary table of haplotypes from RHDdown region represented by population.**

**Table S5C. Summary table of haplotypes from RHCEdown region represented by population.**

	25627957	25683092	25683118	25684012	25685136	25685142	25685448	25685484	25687308	25687426	25687742	25687901	25688146	25688276	25717365	25735202							
	RHD	rs3093653	rs3093654	rs3093632	chr:25685136	rs550214993	rs3093655	rs3093638	rs8876	rs7694	rs3093642	rs1053438	rs148873001	rs9689	rs609320 (E/e)	rs676785 (C/c)	Total	YRI	CHB	CEU	Basques	Catalans	Moroccans
1	T	A	A	T	T	T	G	G	C	A	C	T	A	G	C	G	22.00%	0	0	0	19	3	0
2	C	G	.	C	.	.	.	A	.	.	C	.	A	.	A	18.00%	0	0	0	17	1	0	
3	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	14.00%	0	0	0	13	1	0	
4	C	G	.	C	.	.	.	A	G	.	C	.	A	.	A	6.00%	0	0	0	4	2	0	
5	C	G	.	.	.	.	.	.	.	.	.	.	.	G	.	6.00%	0	0	0	4	2	0	
6	C	G	.	C	.	.	.	.	.	C	.	A	.	A	6.00%	0	0	0	3	3	0		
7	C	G	.	.	.	.	T	.	G	.	.	.	.	G	.	4.00%	0	0	0	4	0	0	
8	C	G	.	C	.	.	.	A	.	.	C	C	A	.	A	3.00%	0	0	0	2	1	0	
9	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	3.00%	0	0	0	2	1	0	
10	.	G	.	C	.	.	.	A	.	.	C	.	A	.	A	3.00%	0	0	0	2	1	0	
11	C	G	.	C	.	.	.	A	.	.	T	C	.	A	.	A	2.00%	0	0	0	2	0	0
12	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	2.00%	0	0	0	2	0	0	
13	.	G	G	.	.	.	.	.	.	.	.	.	.	.	.	2.00%	0	0	0	2	0	0	
14	C	G	.	C	.	.	.	A	.	.	C	.	A	G	.	2.00%	0	0	0	2	0	0	
15	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.00%	0	0	0	1	0	0	
16	.	G	.	.	.	.	T	.	G	.	.	.	.	G	.	1.00%	0	0	0	1	0	0	
17	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	1.00%	0	0	0	1	0	0	
18	C	G	.	C	.	.	.	A	.	.	C	C	A	.	.	1.00%	0	0	0	1	0	0	
19	.	G	.	.	C	.	.	.	.	.	.	.	.	.	.	1.00%	0	0	0	1	0	0	
20	C	G	.	C	.	C	.	A	.	.	C	.	A	.	.	1.00%	0	0	0	1	0	0	
21	.	G	.	C	.	.	.	A	.	.	C	.	A	.	.	1.00%	0	0	0	0	1	0	

**Table S6.**  $F_{ST}$  values of RHD flanking regions calculated by pairwise comparison between populations. P-values are represented above the diagonal, while values of  $F_{ST}$  are showed below.

	RHD upstream $F_{ST}$ \p-value					
	Basques	Catalans	Moroccans	CEU	CHB	YRI
Basques	0	0.727	0.019	0.127	0	0
Catalans	0	0	0.078	0.184	0.015	0
Moroccans	0.040	0.034	0	0.212	0.121	0
CEU	0.013	0.016	0.008	0	0.035	0
CHB	0.089	0.090	0.019	0.040	0	0
YRI	0.277	0.218	0.191	0.255	0.282	0

	RHD downstream $F_{ST}$ \p-value					
	Basques	Catalans	Moroccans	CEU	CHB	YRI
Basques	0	0.871	0.110	0	0	0
Catalans	0	0	0.719	0.003	0.003	0
Moroccans	0.019	0	0	0.000	0	0
CEU	0.091	0.080	0.101	0	0.032	0
CHB	0.097	0.088	0.132	0.037	0	0
YRI	0.141	0.123	0.103	0.110	0.188	0

**Table S7A.**  $F_{ST}$  values of the three RHD/RHCE variants calculated by pairwise comparison between populations for all haplotypes together.

Hudson's $F_{ST}$	Dd (chr1:25627957)	Ee (rs609320 – chr1:25717365)	Cc (rs676785 – chr1:25735202)
BasvsCat	0.000	0.011	0.000
BasvsMor	0.026	0.000	0.000
BasvsCEU	0.013	0.006	0.022
BasvsCHB	0.328	0.082	0.089
BasvsYRI	0.162	0.000	0.381
CatvsMor	0.000	0.019	0.000
CatvsCEU	0.000	0.070	0.016
CatvsCHB	0.250	0.000	0.088
CatvsYRI	0.092	0.044	0.343
MorvsCEU	0.000	0.000	0.000
MorvsCHB	0.159	0.098	0.050
MorvsYRI	0.026	0.000	0.408
CEUvsCHB	0.189	0.161	0.000
CEUvsYRI	0.047	0.000	0.522
CHBvsYRI	0.031	0.130	0.620

**Table S7B.**  $F_{ST}$  values of the three RHD/RHCE variants calculated by pairwise comparison between populations for RhD negative haplotypes.

Hudson's $F_{ST}$	Ee (rs609320 – chr1:25717365)	Cc (rs676785 – chr1:25735202)
BasvsCat	0.000	0.088
BasvsMor	0.000	0.582
BasvsCEU	0.000	0.207
BasvsCHB	0.000	0.000
BasvsYRI	0.000	0.000
CatvsMor	0.000	0.257
CatvsCEU	0.000	0.000
CatvsCHB	0.000	0.000
CatvsYRI	0.000	0.111
MorvsCEU	0.000	0.151
MorvsCHB	0.000	0.000
MorvsYRI	0.000	0.600
CEUvsCHB	0.000	0.000
CEUvsYRI	0.000	0.231
CHBvsYRI	0.000	0.000

**Table S7C.**  $F_{ST}$  values of the three RHD/RHCE variants calculated by pairwise comparison between populations for RhD positive haplotypes.

Hudson's $F_{ST}$	Ee (rs609320 – chr1:25717365)	Cc (rs676785 – chr1:25735202)
BasvsCat	0.0000000	0.0073369
BasvsMor	0.0000000	0.1848956
BasvsCEU	0.0211610	0.0000000
BasvsCHB	0.0075362	0.0000000
BasvsYRI	0.0181818	0.6837618
CatvsMor	0.0095359	0.0053923
CatvsCEU	0.0786207	0.0000000
CatvsCHB	0.0000000	0.0208078
CatvsYRI	0.0717949	0.4862830
MorvsCEU	0.0000000	0.1754739
MorvsCHB	0.0368925	0.1368953
MorvsYRI	0.0000000	0.3033987
CEUvsCHB	0.1074474	0.0000000
CEUvsYRI	0.0000000	0.6763095
CHBvsYRI	0.1004390	0.6385085

## References

- 1 Wagner FF, Flegel W a. RHD gene deletion occurred in the Rhesus box. *Blood* 2000; **95**: 3662–3668.
- 2 McKenna A, Hanna M, Banks E *et al*. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.
- 3 Barrett JC, Fry B, Maller J, Daly MJ. Haplovview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- 4 Gabriel SB, Schaffner SF, Nguyen H *et al*. The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–9.
- 5 Perry GH, Xue Y, Smith RS *et al*. Evolutionary genetics of the human Rh blood group system. *Hum Genet* 2012; **131**: 1205–1216.